Mutations in the *MSH3* gene preferentially lead to deletions within tracts of simple repetitive DNA in *Saccharomyces cerevisiae*

(yeast/microsatellites/DNA mismatch repair)

MICHELINE STRAND*, MARIE C. EARLEY†, GRAY F. CROUSE†, AND THOMAS D. PETES*

*Department of Biology and Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, NC 27599-3280; and †Department of Biology, Emory University, Atlanta, GA 30322

ABSTRACT   Eukaryotic genomes contain tracts of DNA in which a single base or a small number of bases are repeated (microsatellites). Mutations in the yeast DNA mismatch repair genes *MSH2*, *PMS1*, and *MLH1* increase the frequency of mutations for normal DNA sequences and destabilize microsatellites. Mutations of human homologs of *MSH2*, *PMS1*, and *MLH1* also cause microsatellite instability and result in certain types of cancer. We find that a mutation in the yeast gene *MSH3* that does not substantially affect the rate of spontaneous mutations at several loci increases microsatellite instability about 40-fold, preferentially causing deletions. We suggest that *MSH3* has different substrate specificities than the other mismatch repair proteins and that the human *MSH3* homolog (*MRP1*) may be mutated in some tumors with microsatellite instability.

Misincorporation of bases during DNA replication leads to DNA mismatches. In *Escherichia coli*, these mismatches are recognized and corrected by a complex of enzymes including MutS, MutH, and MutL (1). Mutations in the genes encoding these proteins increase the spontaneous mutation rate 100- to 1000-fold. In *Saccharomyces cerevisiae*, genes homologous to *MutS* and *MutL* have been identified. Excluding genes that are meiosis-specific or that affect DNA repair in mitochondria, the MutL homologues are *PMS1* (2) and *MLH1* (3) and the MutS homologues are *MSH2* (4) and *MSH3* (5). Strains with mutations in *MSH2*, *MLH1*, or *PMS1* have about 50-fold elevated spontaneous forward mutation rates (2, 3, 6), whereas the *msh3* mutation has only a 2-fold effect on the mutation rate (5); reversion rates of a frameshift mutation are elevated about 1000-fold by *MSH2*, *MLH1*, or *PMS1* and about 10-fold by *MSH3*. Homologues to the yeast *MSH3* gene exist in mice (7) and humans (8).

Alterations in the length of simple repetitive DNA tracts in yeast are likely to reflect DNA polymerase slippage events (9, 10). In wild-type strains, the displaced repeats resulting from DNA polymerase slippage (Fig. 1) are often corrected by excising the mismatched region from the newly synthesized strand and repairing the gap by using information derived from the template strand. Although repeats are displaced on only one strand in each slippage event, it should be noted that both strands reflect a distortion of the helix. As expected from this model, yeast strains with *pms1*, *msh2*, or *mlh1* mutations have greatly increased levels of instability in microsatellite sequences (9).

Mutations in human homologs of *MSH2*, *MLH1*, and *PMS1* lead to a variety of hereditary tumors (11–15). A diagnostic feature of tumor cells in these patients is unstable microsatellites (16–18). Cell lines derived from some of these tumors are defective in mismatch repair *in vitro* (19, 20). For some sporatic cancers, microsatellite instability has been detected without evidence for a mutation in any of the known mismatch repair genes (21). These tumors may have mutations in unknown mismatch repair genes or in genes encoding proteins affecting DNA replication. Below, we show that mutations in the yeast *MSH3* gene destabilize simple repeats.

EXPERIMENTAL PROCEDURES

Yeast Strains. All strains used in this study were derived from *Saccharomyces cerevisiae* AMY125 (*Matα ade5-1 leu2-3,112 trp1-289 ura 3-52 his 7-2*; obtained from A. Morrison and A. Sugino, Osaka University, Osaka) by transformation. All strains contained either plasmid pSH91 or pSH31 to monitor microsatellite instability (as described below). Strains MS85 and MS96 were Leu+ derivatives of AMY125 obtained by transformation with the *LEU2* integrating plasmid CV9 (9) and contained the assay plasmids pSH31 (MS85) or pSH91 (MS96). MS128 was derived from AMY125 by insertion of an *msh2* deletion allele [constructed by using the plasmid pII-2-Tn10LUK7-7 (6) obtained from E. Alani, Cornell University, Ithaca, NY], transformation with the assay plasmid pSH91, and insertion of a *LEU2* gene by using CV9. MS111 was constructed by transforming AMY125 with CV9 (MS71), transforming MS71 with pEN33 (strain GCY140; contains deletion mutation of *MSH3*), and transforming GCY140 with pSH91. MS125 was constructed by transforming GCY178 with pSH91; GCY178 was made by doing a two-step transplacement of GCY141 (an *msh3* mutant strain isogenic with GCY140) with the plasmid p306M2R1Δ to introduce the *msh2* mutation. MS110 and MS111 are isogenic, as are MS124 and MS125, except for the assay plasmid. MS94 was constructed by transforming AMY125 with the plasmid pII-2-Tn10LUK7-7 to introduce the *msh2* mutation, followed by transformation with the assay plasmid pSH31.

Assays of Microsatellite Instability. Two plasmids, pSH31 (22) and pSH91 (9), were used to measure microsatellite instability. Since the methods used to monitor tract instability have been published (22), the description of these methods is abbreviated. The plasmid pSH31 contained an out-of-frame 29-bp poly(GT) tract within the coding sequence of β-galactosidase (Fig. 2*a*). Strains containing this plasmid formed white colonies on plates containing 5-bromo-4-chloro-3-indolyl β-D-galactoside (X-Gal); strains that had an alteration in the poly(GT) tract length that restored the correct reading frame of the β-galactosidase gene, however, formed blue colonies (as confirmed by DNA sequence analysis). To determine a rate of instability, we determined the frequency of blue colonies in 15–20 cultures. If any of the cultures had no blue

Abbreviations: X-Gal, 5-bromo-4-chloro-3-indolyl β-D-galactoside; 5FOA, 5-fluoroorotate.
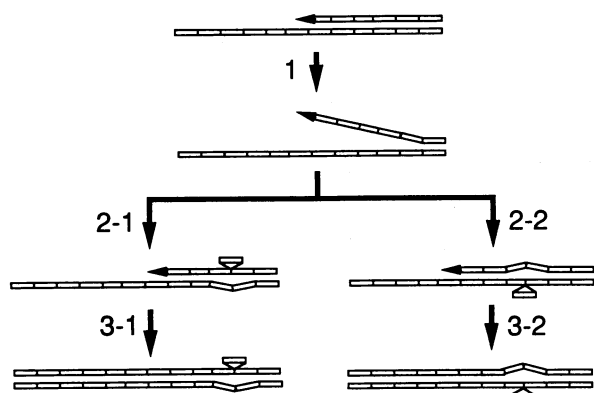
FIG. 1. Alterations in tract length as a consequence of DNA polymerase slippage (9, 10). Only one arm of the replication fork is depicted, with the template strand containing 12 repeats (small rectangles). During replication, the primer and template strand dissociate (arrow 1) and reassociate in a misaligned configuration. If the displaced repeat involves the primer strand (arrow 2-1), continued synthesis (arrow 3-1) results in an increase in tract length. If the displaced repeat involves the template strand (arrow 2-2), then continued synthesis (arrow 3-2) results in a deletion in tract length. Removal of mismatched bases from the primer strand by DNA mismatch repair would restore the wild-type tract length to both strands.

colonies, we used the methods of Luria and Dulbruck (24), as modified by Lea and Coulson (25), to calculate a rate from the frequency data. If all of the cultures had blue colonies, we used the method of the median developed by Lea and Coulson (25). Rates for each strain were determined at least twice.

The plasmid pSH91 had an in-frame 33-bp poly(GT) tract inserted into a fusion protein that had wild-type *URA3* activity (Fig. 2*b*). Thus, yeast strains containing this plasmid were phenotypically Ura$^+$ but failed to grow in medium containing 5-fluoroorotate (5FOA), which selects for Ura$^-$ cells (23). To calculate a rate of instability, we determined the frequency of 5FOA-resistant cells in 15–20 individual cultures. The rate of instability was calculated from these data by the method of the median (25).

Sequence Analysis of Altered Poly(GT) Tracts. As described above, yeast strains containing the plasmid pSH91 were plated onto medium containing 5FOA to detect plasmids with a mutation of the *URA3* gene within the plasmid. Previous experiments indicated that in such plasmids, the alteration involved a change in the length of the poly(GT) tract (22). To determine the types of alterations, we used two different procedures. For all strains except MS96, to measure the size of the poly(GT) tract in these derivatives, we performed the polymerase chain reaction (PCR) procedure on isolated yeast DNA by using primers that flanked the tract (5'-CCAATAG-GTGGTTAGCAATCG and 5'-GTTTTCCCAGTCACGT-AC). The reaction was done with one labeled nucleotide (dATP) and the resulting product was analyzed on a 6% polyacrylamide DNA sequencing gel (26). In all gels, we included standards of PCR products derived from plasmids that had been sequenced that had poly(GT) tracts of 31, 33, and 35 bp. For MS96 (data reported in ref. 1), we rescued plasmids from the 5FOA-resistant yeast strains into *E. coli* DH5α and sequenced the relevant region of the plasmid. In control experiments in which the same plasmids were examined both by the PCR procedure and by plasmid rescue, the same alterations were observed, indicating the validity of both techniques.

Statistical Analysis. Comparisons of the types of sequence alterations observed in different strains were made by using the Fisher exact test (GraphPad INSTAT program on the Macintosh computer). A *P* value of <0.05 was regarded as statistically significant.

RESULTS AND DISCUSSION

Two plasmids were used to examine instability of the simple repeat poly(GT) in yeast (Fig. 2) (9, 22). One plasmid (pSH31) contained an out-of-frame 29-bp insertion of poly(GT) in the coding sequence of an *E. coli* β-galactosidase gene that was fused to a yeast promoter. Yeast cells containing this plasmid failed to make functional β-galactosidase and formed white colonies on plates containing X-Gal. Alterations in tract length that restore the correct reading frame were detected as blue colonies on X-Gal. The second plasmid (pSH91) had an in-frame 33-bp insertion of poly(GT) in the coding sequence of a hybrid gene containing *URA3* sequences. Since cells with a wild-type *URA3* gene are sensitive to the drug 5FOA (23), alterations in the length of the poly(GT) tract can be selected by using medium containing 5FOA.

As reported (9), the rates of microsatellite instability were 3.1×10^{-6} per cell division when measured with plasmid pSH31 and 1.7×10^{-5} per cell division when measured with pSH91 (Table 1). Strains with an *msh2* mutation had greatly elevated levels of instability (9). Depending on which assay is used, the *msh3* mutation resulted in a 25- or 63-fold increase in tract instability compared to wild type. The effect of *msh3* on microsatellite instability was 3- to 9-fold smaller than that observed with *msh2*. The double mutant *msh2 msh3* had a rate
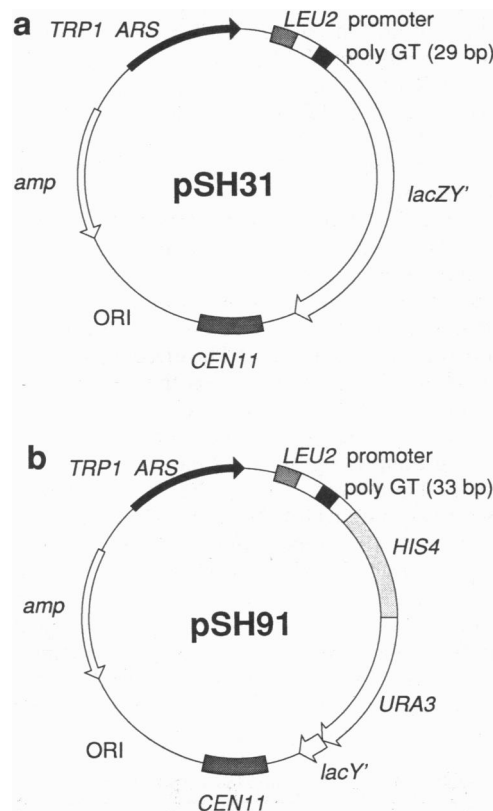

FIG. 2. Plasmids used to assay microsatellite instability. (*a*) pSH31. This plasmid contains the *E. coli* β-galactosidase gene fused to the yeast *LEU2* promoter (22). An out-of-frame 29-bp poly(GT) tract is inserted within the coding sequence of this gene, creating a frameshift mutation. Yeast strains with this plasmid form white colonies on plates containing the chromogenic substrate X-Gal. Strains in which the plasmid-borne poly(GT) tract alters in length to restore the correct reading frame form blue colonies on X-Gal plates. (*b*) pSH91. In this plasmid, the *LEU2* promoter is located upstream of sequences encoding a *HIS4/URA3* fusion protein. An in-frame 33-bp poly(GT) tract is inserted near the beginning of the gene. Strains with pSH91 are phenotypically Ura$^+$. Strains in which the tract alters to cause loss of the correct reading frame can be selected on plates containing the drug 5FOA (23).

Table 1.   Rates of alteration in lengths of poly(GT) tracts in yeast strains with mutations affecting DNA mismatch repair

Strain	Relevant genotype	Tract location	Rate of tract instability relative to wild type (average rate)
MS96	Wild type	pSH91	1
MS128	*msh2*	pSH91	82, 88 (85)
MS111	*msh3*	pSH91	18, 32 (25)
MS125	*msh2 msh3*	pSH91	57, 59, 82, 100 (75)
MS85	Wild type	pSH31	1
MS94	*msh2*	pSH31	303, 742 (523)
MS110	*msh3*	pSH31	39, 87 (63)
MS124	*msh2 msh3*	pSH31	645, 838 (741)

*Numbers on the same line represent independent rate measurements, and numbers in parentheses represent the average rate, relative to a normalized wild-type rate of 1. The absolute rates of tract instability were 1.7×10^{-5} per cell division in MS96 and 3.1×10^{-6} per cell division in MS85.

of instability that is approximately the same as that observed in the single *msh2* mutant.

The types of alterations in the strains containing the pSH91 plasmid were examined (Table 2). In all strains, the most common alterations were additions or deletions of 2 bp. In the wild-type strain, there was a bias in favor of additions; in addition, about 10% of the altered tracts derived from the wild-type strain had large (10 bp or greater) deletions. In both the *msh2* strain and the *msh2 msh3* strain, almost all of the altered tracts represented 2-bp additions or deletions, with a 2-fold bias in favor of deletions. In the *msh3* strain, there was a 6-fold bias in favor of deletions over insertions. This bias was significantly different from that observed for the *msh2* (*P* = 0.02, Fisher exact test) or the *msh2 msh3* (*P* = 0.01, Fisher exact test) strains.

Several lines of evidence indicate Msh3 protein (Msh3p) has a different role from the other mismatch repair proteins. (*i*) Whereas mutations in *msh2, mlh1,* and *pms1* have substantial effects on both spontaneous forward mutation rate and the stability of simple repeats (2, 3, 6, 9), *msh3* substantially affects only simple repeat instability. (*ii*) The effect of *msh3* on simple repeat instability is significantly less than that observed for mutations in the other genes involved in mismatch repair. (*iii*) The ratio of deletions to additions observed in the *msh3* mutant background is significantly more biased toward deletions than that observed in the other mutant backgrounds. Evidence for two systems for the repair of mismatches has also been obtained in *Schizosaccharomyces pombe* (27).

The rates of tract instability observed in the *msh2 msh3* double mutant and the *msh2* single mutant are about the same. Since the rate estimates show a range of 2- to 3-fold in different experiments and since the effect of *msh3* on tract instability is considerably less than that observed for *msh2*, these data do not distinguish whether the effect of *msh3* is epistatic to that of *msh2* or whether the effects of *msh2* and *msh3* are additive.

Table 2.   Sequence analysis of altered poly(GT) tracts

Strain	Relevant genotype	Number of tracts with additions (+) or deletions (−) of base pairs					
		−4	−2	+2	+4	Other	Total
MS96*	Wild type	0	5	23	1	4 (−10, −16, −10, −14)	33
MS128	*msh2*	0	52	22	3	0	77
MS111	*msh3*	2	59	9	2	0	72
MS125	*msh2 msh3*	0	44	23	0	1 (−22)	68

Independent 5FOA-resistant colonies of the above strains were isolated, and changes in the lengths of the poly(GT) tracts in pSH91 plasmids derived from these strains were measured. Numbers in parentheses are the sizes (in bp) of large deletions.
*Data have been described (9).

The data, however, exclude the possibility that the effects are multiplicative as would be expected if the *MSH2* and *MSH3* systems involved different pathways that operated on the same substrate. Mutations in the *MSH2* and *MSH3* genes have additive effects on the frequency of recombination between diverged repeated sequences, indicating that these genes function independently for this recombination pathway (28).

We propose two models consistent with the data. First, the Msh3p may be part of a complex that repairs mismatches independently of the complex involving Msh2 protein (Msh2p). For example, Msh3p could be part of a complex that repairs mismatches on the leading strand during DNA replication and Msh2p part of a complex that repairs mismatches on the lagging strand. By this model, the *msh2 msh3* strain would be expected to have a rate of instability equal to the sum of the rates found in the two single mutant strains. The higher proportion of deletions detected in the *msh3* background could reflect a difference in the types of slippage events observed on the leading and lagging strand or different specificities in the repair of mismatches on the primer or template strands (Fig. 1).

An alternative model is that Msh2p and Msh3p interact with the MutL homologues to generate complexes with different repair efficiencies and specificities. One version of this model is that a heterodimer of Msh2p and Msh3p interacts with the yeast MutL homologues to produce the most efficient repair complex. In the absence of Msh3p, a homodimer of Msh2p interacts with the MutL homologues to produce a complex that is repair-proficient for single base-pair mismatches but less efficient in the repair of mismatches formed by 2-bp loops. In the absence of Msh2p, no repair complex is formed or the complex formed by a homodimer of Msh3p interacting with the MutL homologues functions poorly for all types of mismatch repair. To explain the bias in favor of deletions observed in the *msh3* mutant strain, we would also have to postulate that the two repair complexes have different efficiencies for repairing loops formed on the template and primer strands.

Prolla *et al.* (29) suggested that Msh2p, Pms1 protein, and Mlh1 protein can interact in formation of a ternary complex on DNA. It would be of interest to determine whether the addition of Msh3p would allow formation of a quaternary complex. Since there are a number of MutL and MutS homologues in addition to those described above, other repair complexes involving various MutL- and MutS-like proteins are also possible.

As summarized above, *msh3* mutants have large effects on microsatellite instability and small effects in a forward mutation assay. In mammalian cells, mutations in the p160 subunit of the hMutSα heterodimer have the opposite effect: elevating the rate of mutation in a forward assay at the *HPRT* locus without a comparable increase in the rate of microsatellite instability (30). Thus, in both yeast and mammalian cells, the repair of different types of DNA mismatches is differentially sensitive to mutations affecting different components of the mismatch repair system.

Our results suggest that the human homologue of *MSH3* (8) is a candidate gene for mutation in tumors that show microsatellite instability but no alterations in other mismatch repair genes. Tumors in which the simple repeats are preferentially deleted would be of particular interest. Since the *msh3* mutation primarily affects frameshift mutations, it will have a cancer-producing phenotype only if the target tumor suppressor genes contain regions with repeated bases. Markowitz *et al.* (31) have reported that one target gene in cancer cells with microsatellite instability is a gene encoding a type II transforming growth factor *β* receptor and that mutations occur in a simple repeat within this gene.

1. Modrich, P. (1991) *Annu. Rev. Genet.* **25,** 220–253.
2. Kramer, W., Kramer, B., Williamson, M. S. & Fogel, S. (1989) *J. Bacteriol.* **171,** 5339–5346.
3. Prolla, T. A., Christie, D. M. & Liskay, R. M. (1994) *Mol. Cell Biol.* **14,** 407–415.
4. Reenan, R. A. G. & Kolodner, R. D. (1992) *Genetics* **132,** 963–973.
5. New, L., Lui, K. & Crouse, G. F. (1993) *Mol. Gen. Genet.* **239,** 97–108.
6. Reenan, R. A. G. & Kolodner, R. D. (1992) *Genetics* **132,** 975–985.
7. Linton, J. P., Yen, J.-Y. J., Selby, E., Chen, Z., Chinsky, J. M., Liu, K., Kellems, R. E. & Crouse, G. F. (1989) *Mol. Cell. Biol.* **9,** 3058–3072.
8. Shinya, E. & Shimada, T. (1994) *Nucleic Acids Res.* **22,** 2143–2149.
9. Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. (1993) *Nature (London)* **365,** 274–276.
10. Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31,** 77–84.
11. Fishel, R., Lescoe, M. K., Rao, M. R. S., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M. & Kolodner, R. (1993) *Cell* **75,** 1027–1038.
12. Leach, F., Nicolaides, N. C., Papadopoulos, N., Liu B., Jen, J. *et al.* (1993) *Cell* **75,** 1215–1225.
13. Bronner, C. E., Baker, S. M., Morrison, P. T., Warren, G., Smith, L., Lescoe, M. K., Kane, M., Earabino, C., Lipford, J., Lindblom, A., Tannergard, P., Bollag, R. J., Godwin, A. R., Ward, D. C., Nordenskjold, M., Fishel, R., Kolodner, R. & Liskay, R. M. (1994) *Nature (London)* **368,** 258–261.
14. Nicolaides, N. C., Papadopoulos, N., Lui, B., Wei, Y., Carter, K. C., Ruben, S. M., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., Adams, M. D., Venter, J. C., Dunlop, M. G., Hamilton, S. R., Petersen, G. M., de la Chapelle, A., Vogelstein, B. & Kinzler, K. (1994) *Nature (London)* **371,** 75–80.
15. Papadopoulos, N., Nicolaides, N. C., Wei, Y.-F., Ruben, S. M., Carter, K. C., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., Adams, M., Venter, J. C., Hamilton, S., Petersen, G. M., Watson, P., Lynch, H. T., Peltomaki, P., Mecklin, J.-P., de la Chapelle, A., Kinzler, K. W. & Vogelstein, B. (1994) *Science* **263,** 1625–1629.
16. Aaltonen, L. A., Peltomaki, P., Leach, F. S., Sistonen, P., Pylkkanen, L., Mecklin, J. P., Jarvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., Peterson, G. N., Kinzler, K. W., Vogelstein, B. & de la Chapelle, A. (1993) *Science* **260,** 812–818.
17. Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M. (1993) *Nature (London)* **263,** 558–561.
18. Thibodeau, S. N., Bren, G. & Schaid, D. (1993) *Science* **260,** 818–821.
19. Parsons, R., Li, G.-M., Longley, M. J., Fang, W.-h., Papadopoulos, N., Jen, J., de la Chapelle, A., Kinzler, K. W., Vogelstein, B. & Modrich, P. (1993) *Cell* **75,** 1227–1236.
20. Umar, A., Boyer, J. C., Thomas, D. C., Nguyen, D. C., Risinger, J. I., Boyd, J., Ionov, Y., Perucho, M. & Kunkel, T. A. (1994) *J. Biol. Chem.* **269,** 14367–14370.
21. Liu, B., Nicolaides, N. C., Markowitz, S., Willson, J. K. V., Parsons, R. E., Jen, J., Papadopolous, N., Peltomaki, P., de la Chapelle, A., Hamilton, S. R., Kinzler, K. W. & Vogelstein, B. (1995) *Nat. Genet.* **9,** 48–55.
22. Henderson, S. T. & Petes, T. D. (1992) *Mol. Cell. Biol.* **12,** 2749–2757.
23. Boeke, J. D., Lacroute, F. & Fink, G. (1984) *Mol. Gen. Genet.* **197,** 345–346.
24. Luria, S. E. & Delbruck, M. (1943) *Genetics* **28,** 491–511.
25. Lea, D. E. & Coulson, C. A. (1949) *J. Genet.* **49,** 264–285.
26. Farber, R. A., Petes, T. D., Dominska, M., Hudgens, S. S. & Liskay, R. M. (1994) *Hum. Mol. Genet.* **3,** 253–256.
27. Schar, P. & Kohli, J. (1993) *Genetics* **133,** 825–835.
28. Selva, E. M., New, L., Crouse, G. F. & Lahue, R. S. (1995) *Genetics* **139,** 1175–1188.
29. Prolla, T. S., Pang, Q., Alani, E., Kolodner, R. D. & Liskay, R. M. (1994) *Science* **265,** 1091–1093.
30. Drummond, J. T., Li, G.-m., Longley, M. J. & Modrich, P. (1995) *Science* **268,** 1909–1912.
31. Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R. S., Zborowska, E., Kinzler, K. W., Vogelstein, B., Brattain, M. & Willson, J. K. V. (1995) *Science* **268,** 1336–1338.